

**This paper was published in *Journal of the Optical Society of America A* and is made available as an electronic reprint with the permission of OSA. The paper can be found at the following URL on the OSA website:**

**<http://www.opticsinfobase.org/josaa/abstract.cfm?URI=josaa-29-2-A194>.**

**Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.**

# Visual search in natural scenes explained by local color properties

Kinjiro Amano,\* David H. Foster, Matthew S. Mould, and John P. Oakley

*School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, UK*

*\*Corresponding author: k.amano@manchester.ac.uk*

Received September 1, 2011; revised November 17, 2011; accepted November 18, 2011;  
posted November 21, 2011 (Doc. ID 153766); published January 26, 2012

Success in visually searching for a small object or target in a natural scene depends on many factors, including the spatial structure of the scene and the pattern of observers' eye movements. The aim of this study was to determine to what extent local color properties of natural scenes can account for target-detection performance. A computer-controlled high-resolution color monitor was used to present images of natural scenes containing a small, randomly located, shaded gray sphere, which served as the target. Observers' gaze position was simultaneously monitored with an infrared video eye-tracker. About 60% of the adjusted variance in observers' detection performance was accounted for by local color properties, namely, lightness and the red-green and blue-yellow components of chroma. A similar level of variance was accounted for by observers' fixations. These results suggest that local color can be as influential as gaze position in determining observers' search performance in natural scenes. © 2012 Optical Society of America

OCIS codes: 330.1720, 330.2210, 330.1880.

## 1. INTRODUCTION

Looking for an object in the world around us is a common part of everyday experience [1,2]. Success or otherwise in this task depends on many factors, including the structure of the environment or scene being searched, the nature of the object being sought, and the task itself [3–8]. Evidence for the contribution of physical factors in determining visual search, such as the spatial-frequency and orientation content of the scene and of the object, has, however, come largely from experiments with simple abstract geometric patterns containing a target in a background of distractor elements [9–11]. Color has been found to be important in determining search in such experiments, but primarily as a way of defining differences between the target and the background [9,10,12–14]. Rather less is known about visual search in natural scenes, with most work concentrating on the separate question of eye movements and visual saliency [2,7,15–17].

Studies of free viewing, i.e., not involving search, have concentrated on the role of achromatic spatial contrast [18–20]. One such analysis [21] with natural scenes suggested that only edge density was a significant predictor of observers' fixations, compared with other gray-level image properties such as extrema in contrast and in luminance and high spatial-frequency content. Another study [22] concluded that higher-order statistical properties associated with a bispectral analysis was a significant factor for saccadic selection. More generally, local features in gray-level images have been deduced [23] to account at best for about 60% of the variance in point of gaze.

Yet color does seem to be relevant. Again with free viewing, different patterns of fixations have been reported [24,25] with colored and gray-level versions of the same natural scenes. Nevertheless, the influence of local color properties on search performance remains unclear [cf., 26,27]. The aim of this study, therefore, was to determine explicitly to what extent

local color properties of natural scenes can account for performance in search for and detection of a target.

To this end, observers were presented with images of natural scenes rendered on a color monitor. The target was a small, shaded, gray sphere embedded in the scene and matched in mean luminance to its local surround to avoid producing accidental chromatic or luminance contrast cues to detection. To provide an alternative spatial account of detection performance [cf., 23], observers' gaze position was simultaneously monitored with an infrared video eye-tracker. The accuracy of predictions based on the spatial distribution of local color properties was then compared with those based on the spatial distribution of observers' fixations.

It was found that 57%–60% of the variance in observers' detection performance over scenes could be explained by local color properties, namely, lightness, and the red-green and blue-yellow components of chroma, a level closely similar to that explained by the distribution of observers' fixations. Despite a general assumption that spatial factors such as contrast and edges dominate visual search and gaze behavior [e.g., 18–20,21,22], the present results suggest that local color information can be as important as gaze position in determining observers' ability to search for and detect a target in natural scenes.

## 2. METHODS

### A. Apparatus

Images of natural scenes were presented on the screen of a 20-inch RGB CRT color monitor (GDM-F520, Sony Corp., Tokyo, Japan) controlled by a graphics system (Fuel, Silicon Graphics Inc., California, USA) with spatial resolution 1600 × 1200 pixels, refresh rate approximately 60 Hz, and intensity resolution 10 bits on each R, G, and B channel. The colorimetric accuracy of the display system was regularly calibrated and tested with a telespectroradiometer (SpectraColorimeter

PR-650, Photo Research Inc., Chatsworth, California, USA) and photometer (LMT, L1003, Lichtmesstechnik GmbH, Berlin, Germany) whose accuracy was verified against other calibrated instruments. Errors in a colored test patch were  $<0.005$  in CIE 1931  $(x, y)$  chromaticity coordinates and  $<5\%$  in luminance.

Observers' eye-movements were recorded with an infrared monocular video eye-tracker (High Speed Video Eyetracker Toolbox mk2, Cambridge Research Systems Ltd., Kent, UK), with a temporal sampling frequency of 250 Hz, connected to a computer that recorded and analyzed the output signals.

## B. Stimuli

Twenty natural scenes were taken from a set of hyperspectral images [28] to allow accurate calculation of the local colorimetric properties under the chosen scene illumination, an average daylight of correlated color temperature 6500 K [29]. The images rendered on the monitor screen subtended approximately  $17 \times 13$  deg of visual angle at the viewing distance of 1 m. The mean luminance of the images on the screen was  $3.6 \text{ cd m}^{-2}$  (range  $0\text{--}61.4 \text{ cd m}^{-2}$ ). Figure 1 (top row) shows some example images. The target is indicated by the arrow in the rightmost image in Fig. 1, top row, with a close-up in the inset. Its surface was spectrally neutral (Munsell N7) and it subtended approximately 0.25 deg of visual angle in each of the 20 scenes. It appeared randomly in each image at one of 130 possible locations, defined by an imaginary  $13 \times 10$  grid. The illumination and shading on the sphere were consistent with that on the scene since the sphere was embedded in the scene at the time of hyperspectral imaging. As already mentioned, trivial brightness cues were eliminated by matching the average luminance of the target at each loca-

tion to the average luminance of its local surround ( $<1.0$  deg extent). Because of their shading, targets were not always isoluminant with their local surrounds. The relatively small angular subtense of the target was chosen, as a result of preliminary measurements, to encourage observers to inspect the entire image and to avoid a strategy of "the best search is no search at all" [30]. The difficulty of finding the target without search should be evident from the example of Fig. 1, top right.

## C. Procedure

In each trial, observers were presented with an image of a particular scene for 1 s, followed by a dark field. They had to indicate whether or not they saw the target by pressing a computer mouse button (a "yes/no" task). They were allowed to move their eyes during the trial and had unlimited time to respond. The next trial started approximately 0.7 s after that response. No fixation point was presented either before or during the trial to guide observers' gaze position, which was recorded continuously. Head movement was minimized with a forehead rest and chinrest.

Each scene was tested in 260 trials, which constituted one experimental block. Half of the trials, chosen at random, contained the target, and the other half did not. Experimental blocks were divided into four subblocks of 65 trials. Observers performed a calibration of the eye-tracker at the start, in the middle, and at the end of each subblock, and they were allowed to take a short break between subblocks. In total each observer performed 5200 trials (20 scenes  $\times$  260 trials), with additional trials repeated if there was a failure with the eye-tracker.

In each calibration of the eye-tracker, observers were presented with 20 calibration targets, consisting of black 0.25-deg

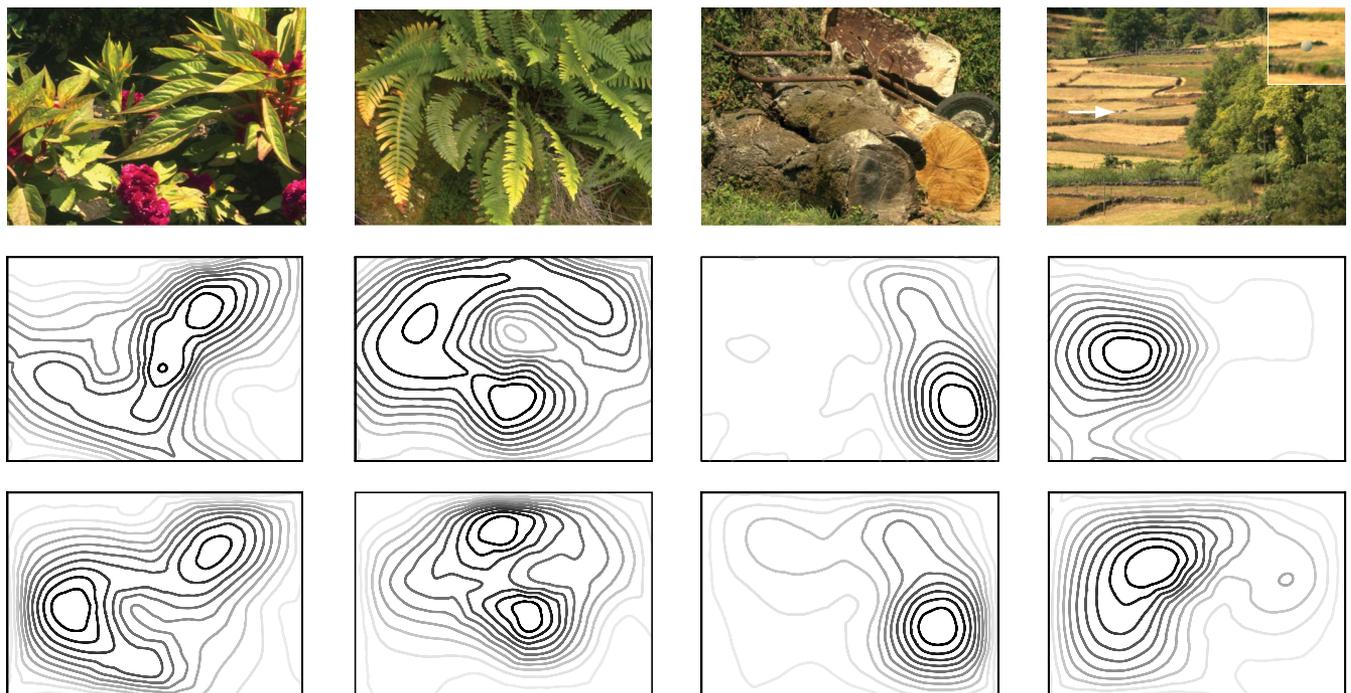


Fig. 1. (Color online) Examples of images, observers' detection performance, and eye fixations. The top row shows four of the natural scenes taken from the 20 used in the experiment. The location of the target in the rightmost image is indicated by the arrow, and the target itself is shown in the inset. The middle row shows the spatial distributions of observers' smoothed detection performance  $d'$  for the corresponding scenes in the top row. Higher values of  $d'$  are indicated by darker contours. The bottom row shows the spatial distribution of smoothed fixations for each of the corresponding scenes in the top row. Higher densities of fixations are indicated by darker contours. The loess smoothing bandwidth was 0.15.

crosses arranged in a  $5 \times 4$  grid on the screen at known positions, each of which they fixated in turn, signaling their fixation with a mouse button. Each of the 20 gaze positions was averaged over the start and middle calibrations, and over the middle and end calibrations. An affine transformation was applied to these mean positions to fit the known screen coordinates to give least-squares error. The resulting coefficients of the transformation were used to transform the experimental gaze-positions in each subblock to screen coordinates.

The root-mean-square error between the 20 calibration targets and the observer's corresponding gaze-positions was taken as the calibration error. Over observers and scenes, based on 1120 measurements, the mean calibration error was approximately 0.26 deg [standard deviation (SD) 0.06 deg].

#### D. Observers

Seven observers (four female, three male, aged 21–31 years) took part in the experiment. All of them had normal binocular visual acuity and normal color vision, verified with a series of color vision tests (Farnsworth-Munsell 100-Hue test, Ishihara pseudoisochromatic plates, Rayleigh and Moreland anomaloscopy with luminance test). All of the observers except one (coauthor MSM) were unaware of the purpose of the experiment. The procedure was approved by the University of Manchester Committee on the Ethics of Research on Human Beings, which operated in accord with the principles of the Declaration of Helsinki.

### 3. RESULTS

#### A. Target-Detection Performance

Detection performance at each of the 130 target locations was quantified by the discrimination index  $d'$  from signal-detection theory [31], where if  $\Phi$  is the normal cumulative distribution function, HR is the hit rate, and FAR is the false-alarm rate, then  $d' = \Phi^{-1}(\text{HR}) - \Phi^{-1}(\text{FAR})$ . The HR is the relative frequency with which the target is reported as being present when it is present, and the FAR is the relative frequency with which the target is reported as being present when it is not. The HR was recorded for each of the 130 target locations within each scene. Since the position of the target is not defined when the target is absent, the FAR is a property of the scene itself rather than of any particular location. For the purpose of calculating  $d'$ , therefore, the FAR was treated as being uniform over each scene, although this assumption has no physical correlate. The value of this FAR varied from scene to scene.

Values of  $d'$  for individual observers' performance averaged over the 20 scenes ranged from 0.7 to 1.8 (mean 1.18, SD 0.34), but trends over scenes were coherent (e.g., all observers performed less well with the same scenes). A one-way repeated-measures ANOVA confirmed a significant effect of scenes ( $F(19, 114) = 15.8, p < 0.05$ ).

Because only one target-detection response was available from each of the seven observers at each of the 130 locations in each of the 20 scenes, it was necessary to pool responses to estimate the raw HR. Even so, these estimates at each location in each scene were based on only seven trials. To improve the accuracy of the  $d'$  estimates at each of the 130 locations and to facilitate comparisons with other data (i.e., point-of-gaze data and local color properties), the distribution of raw  $d'$  values

over each scene was smoothed by a locally weighted quadratic polynomial regression, loess [32,33]. The bandwidth of the loess smooth was varied in 0.05 steps over the range 0.15–0.3, although the precise choice had little effect on the regressions considered later (Subsection 3.D). Figure 1, middle row, shows the smoothed spatial distribution of  $d'$  values for the four example scenes of Fig. 1, top row. Higher values of  $d'$  are indicated by darker contours.

#### B. Point of Gaze

There were several algorithms available for classifying fixations from point-of-gaze data [27,34,35], and the method used here was taken from [36], which had been found to produce classifications similar to those based on a nonparametric method [37]. Saccades were demarcated by a velocity-based, two-step procedure, in which saccade detection was initiated when the gaze velocity first exceeded a higher threshold, and then the onset and offset of the saccade were set, respectively, by the first samples where the velocity rose above and fell below a lower threshold. The two velocity thresholds were adjusted for individual observers [38]. These higher and lower thresholds ranged, respectively, over 94–140 deg s<sup>-1</sup> and 20–30 deg s<sup>-1</sup> for the seven observers. This variation in thresholds over observers is not unusual [38]. Intervals between saccades were assumed to be fixations. All fixations up to the fourth after the first saccade were included in the analysis of each trial. In all, approximately 95,000 fixations were used.

For the same reasons that applied in estimating  $d'$ , the distributions of fixations over each scene were pooled over observers and then smoothed by the same locally weighted quadratic regression as in Subsection 3.A. Figure 1, bottom row, shows the smoothed spatial distribution of fixations for the four example scenes of Fig. 1, top row. Higher densities of fixations are indicated by darker contours.

The relationship between the spatial distribution of fixations and  $d'$  values is considered in Subsection 3.D.

#### C. Local Color Properties

The local color properties of each scene were quantified by the CIE color appearance model CIECAM02 [39–41]. But it is emphasized that this choice was based on computational convenience, deriving from the approximate perceptual uniformity of CIECAM02, rather than for any particular appearance attributes. The color space CIELAB [29] could have been used instead, but it would have been necessary to incorporate the color-difference formula CIEDE2000 [29,42] in the spatial smoothing operation.

CIECAM02 provides among other quantities lightness ( $J$ ), chroma ( $C$ ), and hue ( $h$ ) in a polar coordinate system, or, equivalently, lightness ( $J$ ) and the red-green chroma component ( $a_C$ ) and blue-yellow chroma component ( $b_C$ ) in a Cartesian coordinate system. The parameters [41] of the CIECAM02 model were set so that the white point was D65, the luminance of the background 20% of the white level, and the surround “average.”

The distributions of these three quantities  $J$ ,  $a_C$ , and  $b_C$  over each scene were smoothed by the same locally weighted quadratic regression as in Subsections 3.A and 3.B. Figure 2 shows the smoothed spatial distributions of  $J$ ,  $a_C$ , and  $b_C$  for the four example scenes of Fig. 1. Higher values are indicated by darker contours.

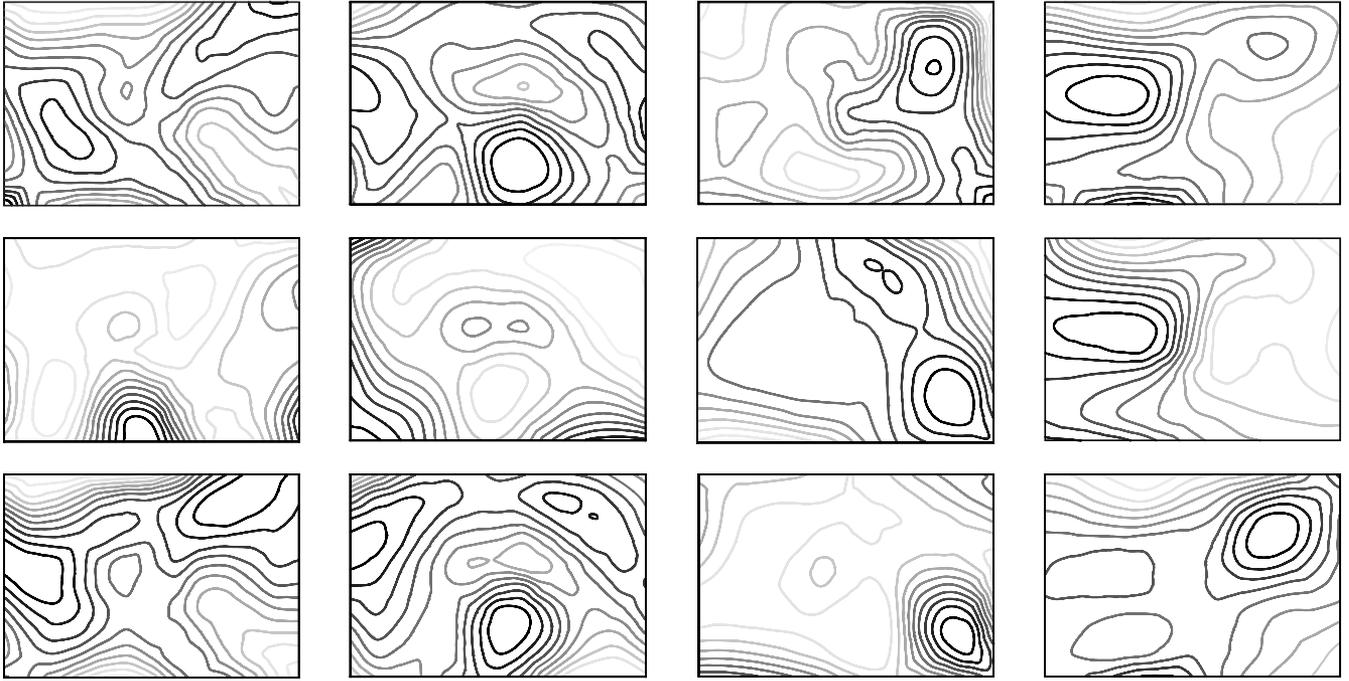


Fig. 2. Spatial distributions of color properties for each of the corresponding scenes in Fig. 1. The top row shows the smoothed lightness  $J$ ; the middle row, the smoothed red-green chroma component  $a_C$ ; and the bottom row, the smoothed blue-yellow chroma component  $b_C$ . Higher values are indicated by darker contours. The loess smoothing bandwidth was 0.15.

The relationships between the spatial distributions of  $d'$  values, observers' fixations, and local color properties are considered next.

#### D. Regression Analysis

To assess the dependence of target-detection performance on local color properties, smoothed values of the discrimination index  $d'$  (as in Fig. 1, middle row) were regressed on smoothed  $J$ ,  $a_C$ , and  $b_C$  values (as in Fig. 2, top, middle, and bottom rows, respectively) in each scene. Thus, if  $\hat{d}'(x, y)$ ,  $\hat{J}(x, y)$ ,  $\hat{a}_C(x, y)$ , and  $\hat{b}_C(x, y)$  represent the smoothed values of  $d'$ ,  $J$ ,  $a_C$ , and  $b_C$ , respectively, at location  $(x, y)$ , then the full regression equation had the following form:

$$E[\hat{d}'(x, y)] = \beta_1 \hat{J}(x, y) + \beta_2 \hat{a}_C(x, y) + \beta_3 \hat{b}_C(x, y) + \alpha, \quad (1)$$

where  $E$  is the expectation, and  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , and  $\alpha$  are scalars. Goodness of fit was summarized by  $R^2$ , the proportion of variance accounted for. The value of  $R^2$  was adjusted for the loss in degrees of freedom (d.f.) in smoothing [33] and in fitting the regressor variables [43]. Thus, the adjusted value of  $R^2$  was defined as  $1 - (1 - R^2)(n - 1)/(n - k)$ , where  $n$  is the d.f. of the smooth [33] and  $k = 2$  for the regression on, e.g.,  $J$  and  $k = 4$  for the regression on  $J$ ,  $a_C$ , and  $b_C$ . For comparison, the same analysis was performed with smoothed  $d'$  values regressed on the smoothed distribution of fixation positions in each scene.

Table 1 lists the adjusted  $R^2$  values, averaged over the 20 scenes, along with means and standard errors (SEMs), for each of the color variables and their combination, and for fixation position.

The three color properties  $J$ ,  $a_C$ , and  $b_C$  each accounted for about the same amount of variance in detection performance, namely, 29–35%, over the four values of the smoothing

bandwidth. Combining  $J$  with either  $a_C$  or  $b_C$  increased  $R^2$  to 45–49%. But all three properties together accounted for 57–60% of the variance, almost the same as that accounted for by fixation position, namely, 53–60%. In fact, there was no significant difference over scenes between the explanatory power of the three color properties and fixation position ( $t(34) \leq 0.54$ ,  $p \geq 0.5$ ).

Unsurprisingly, the sum of the  $R^2$  values from each of the color properties was less than when they were combined, suggesting some dependence between the properties in their contribution to target-detection performance.

For comparison, when the smoothed distribution of observers' fixations was regressed on smoothed  $J$ ,  $a_C$ , and  $b_C$  values, exactly as in Eq. (1) for  $d'$ , it was found again that each accounted for about the same amount of variance in detection performance, namely, 17–22%, and all three properties together accounted for 36–40% (not shown in Table 1). These values of  $R^2$  were lower than the corresponding values for target-detection performance, but not significantly so ( $t(37) \geq 1.2$ ,  $p \leq 0.25$ ).

## 4. DISCUSSION

The factors determining visual search in natural scenes are relatively poorly understood, but it has been generally assumed that local color properties are not strong determiners of performance [24,25,44–46]. As has been shown here, however, local values of lightness and the two chroma components can explain 57–60% of the variance in observers' detection performance, averaged over scenes. This level is high for any scene attribute, given that optimal nonparametric models of eye movements in free viewing of natural scenes have difficulty in explaining more than about 60% of performance [23]. The fact that the proportion of variance accounted for varied little with different values of the spatial

**Table 1. Accounting for Variance in Detection Performance<sup>a</sup>**

	Bandwidth 0.15		Bandwidth 0.30	
	Mean <sup>b</sup> %	SEM <sup>b</sup> %	Mean <sup>b</sup> %	SEM <sup>b</sup> %
Lightness $J$	31.0	5.4	33.3	5.6
Chroma projection $a_C$	28.8	6.0	29.4	6.0
Chroma projection $b_C$	34.2	5.4	34.5	5.0
Combined $J + a_C$	46.4	5.9	49.1	6.0
Combined $J + b_C$	44.7	5.8	46.1	5.8
Combined $J + a_C + b_C$	56.9	4.8	60.3	4.0
Fixation position	53.1	5.3	60.3	5.5

<sup>a</sup>Mean values of  $R^2$  are shown for regressions of the smoothed discrimination index  $d'$  on smoothed individual local color properties in the color space CIECAM02 [41], on their smoothed combinations, and on the smoothed distribution of fixation positions. Discrimination performance and fixations were each based on pooled data from observers. Values of  $R^2$  were adjusted for loss in degrees of freedom in smoothing [33] and in fitting [43]. Values of the loess smoothing bandwidth are indicated.

<sup>b</sup>Means and SEMs were obtained over 20 natural scenes.

smoothing parameter suggests that this result is not due to an accident of over- or under-smoothing.

Although the distribution of eye movements is known to be different in search and in free viewing [16,27,47,48], it is revealing in this analysis that observers' fixations explained 53–60% of the variance in detection performance, again averaged over scenes, not very different from the maximum estimated for eye movements in free viewing [23]. Local color information thus seems to be as effective as gaze position in influencing observers' detection performance.

A prerequisite for the present study was a wide range of color variation across the stimulus images. The natural scenes used here included the main vegetated and nonvegetated land-cover classes, namely woodland, vegetation (e.g., grasses, ferns, flowers), cultivated land, and urban residential and commercial buildings. The variety of color properties is illustrated by the four example images of Fig. 1, top row, where lightness  $J$  ranged over 3.7–35.1, chroma  $C$  over 3.3–39.2, and hue  $h$  over 5.8–152.6 deg. More detailed analysis of these color-image statistics has been reported elsewhere [49].

As might be expected from this variation in the composition of natural scenes, the slopes of the regressions of detection performance on local color properties varied from scene to scene. In particular, the slope for lightness  $J$  was little different from zero when averaged over scenes, although the averaged slopes for the two chroma components were positive. The implication of this is that the predictive capacity of local color properties may be specific to individual scenes, or classes of scenes. In other words, if detection performance is known in one part of a scene, then it is possible to predict detection performance in another part of that scene, but not necessarily in other scenes.

One potential confound in describing observers' gaze behavior is a well-established central bias with circumscribed images of scenes. Thus, rather than being distributed uniformly, fixations tend to be directed toward the center of the image, and it has been argued that the nature of the experimental task and the images themselves may facilitate such a bias [21,27,46]. Yet it seems unlikely that such a bias could account for the distributions of fixations and of target-detection performance found in this study. As is evident in Fig. 1, bottom row, the distributions were markedly scene-dependent.

Target color and scene color are, of course, influential in a range of tasks involving natural scenes, for example, in discriminating fruit and fresh foliage from more mature foliage [50,51], in memorizing and recognizing scenes [52,53], and in extracting the gist of a scene [54]. All these activities, however, are associated with specific tasks and goals, which affect the observed pattern of visual performance. The use here of a neutral target matched in average luminance to its local surround allowed the low-level properties of natural scenes themselves to be probed, thereby revealing the significant influence of color on detection performance. The nature of this influence and of achromatic scene features mentioned earlier presumably reflects the kind of representation afforded to scenes during the early stages of vision. Such a representation might provide the default basis for visual search, on which the requirements of more specific tasks and goals could be superimposed.

## ACKNOWLEDGMENTS

We thank G. Feng for critically reading the manuscript. This work was supported by the Engineering and Physical Sciences Research Council (grant EP/F023669/1). A preliminary report of the results was presented at the 21st Symposium of the International Colour Vision Society, Kongsberg, Norway, July 2011.

## REFERENCES

1. J. M. Wolfe, M. L.-H. Võ, K. K. Evans, and M. R. Greene, "Visual search in scenes involves selective and nonselective pathways," *Trends Cogn. Sci.* **15**, 77–84 (2011).
2. R. van der Lans, R. Pieters, and M. Wedel, "Eye-movement analysis of search effectiveness," *J. Am. Stat. Assoc.* **103**, 452–461 (2008).
3. M. J. Bravo and K. Nakayama, "The role of attention in different visual-search tasks," *Percept. Psychophys.* **51**, 465–472 (1992).
4. J. M. Wolfe, "Visual search in continuous, naturalistic stimuli," *Vision Res.* **34**, 1187–1195 (1994).
5. X. Chen and G. J. Zelinsky, "Real-world visual search is dominated by top-down guidance," *Vision Res.* **46**, 4118–4133 (2006).
6. J. M. Henderson, G. L. Malcolm, and C. Schandl, "Searching in the dark: cognitive relevance drives attention in real-world scenes," *Psychon. B. Rev.* **16**, 850–856 (2009).
7. A. Torralba, A. Oliva, M. S. Castelano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search," *Psychol. Rev.* **113**, 766–786 (2006).
8. P. G. Lovell, I. D. Gilchrist, D. J. Tolhurst, and T. Troscianko, "Search for gross illumination discrepancies in images of natural objects," *J. Vision* **9**(1), 37 (2009).
9. A. Treisman and S. Sato, "Conjunction search revisited," *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 459–478 (1990).
10. A. L. Nagy and G. Thomas, "Distractor heterogeneity, attention, and color in visual search," *Vision Res.* **43**, 1541–1552 (2003).
11. D. H. Foster and P. A. Ward, "Asymmetries in oriented-line detection indicate two orthogonal filters in early vision," *Proc. R. Soc. B* **243**, 75–81 (1991).
12. J. Duncan and G. W. Humphreys, "Visual search and stimulus similarity," *Psychol. Rev.* **96**, 433–458 (1989).
13. D. T. Lindsey, A. M. Brown, E. Reijnen, A. N. Rich, Y. I. Kuzmova, and J. M. Wolfe, "Color channels, not color appearance or color categories, guide visual search for desaturated color targets," *Psychol. Sci.* **21**, 1208–1214 (2010).
14. B. L. Cole and K.-Y. Lian, "Search for coloured objects in natural surroundings by people with abnormal colour vision," *Clin. Exp. Optom.* **89**, 144–149 (2006).
15. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).

16. T. Foulsham and G. Underwood, "What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition," *J. Vision* **8**(2), 6 (2008).
17. G. J. Zelinsky, W. Zhang, B. Yu, X. Chen, and D. Samaras, "The role of top-down and bottom-up processes in guiding eye movements during visual search," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, eds. (MIT Press, 2006), pp. 1569–1576.
18. P. Reinagel and A. M. Zador, "Natural scene statistics at the centre of gaze," *Netw. Comput. Neural Syst.* **10**, 341–350 (1999).
19. D. J. Parkhurst and E. Niebur, "Scene content selected by active vision," *Spatial Vis.* **16**, 125–154 (2003).
20. W. Einhäuser and P. König, "Does luminance-contrast contribute to a saliency map for overt visual attention?," *Eur. J. Neurosci.* **17**, 1089–1097 (2003).
21. S. K. Mannan, K. H. Ruddock, and D. S. Wooding, "The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images," *Spatial Vis.* **10**, 165–188 (1996).
22. G. Krieger, I. Rentschler, G. Hauske, K. Schill, and C. Zetsche, "Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics," *Spatial Vis.* **13**, 201–214 (2000).
23. W. Kienzle, F. A. Wichmann, B. Schölkopf, and M. O. Franz, "A nonparametric approach to bottom-up visual saliency," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, eds. (MIT Press, 2007), pp. 689–696.
24. H.-P. Frey, C. Honey, and P. König, "What's color got to do with it? The influence of color on visual attention in different categories," *J. Vision* **8**(14), 6 (2008).
25. H.-P. Frey, K. Wirz, V. Willenbockel, T. Betz, C. Schreiber, T. Troscianko, and P. König, "Beyond correlation: do color features influence attention in rainforest?," *Front. Hum. Neurosci.* **5**, 36 (2011).
26. W. Einhäuser, U. Rutishauser, and C. Koch, "Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli," *J. Vision* **8**(2), 2 (2008).
27. B. W. Tatler, "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *J. Vision* **7**(14), 4 (2007).
28. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *J. Opt. Soc. Am. A* **23**, 2359–2372 (2006).
29. CIE, Technical Committee 1-48, "Colorimetry," 3rd ed., CIE 15:2004 (Commission Internationale de l'Éclairage, 2004).
30. H. Corbin, J. Carter, E. P. Reese, and J. Volkman, *Experiments on Visual Search 1956–1957* (Psychological Research Unit, Mount Holyoke College, 1958).
31. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, 1966).
32. W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
33. T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models, CRC Monographs on Statistics & Applied Probability* (Chapman & Hall/CRC, 1990).
34. M. Böhme, M. Dorr, C. Krause, T. Martinetz, and E. Barth, "Eye movement predictions on natural videos," *Neurocomput.* **69**, 1996–2004 (2006).
35. M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behav. Res. Methods* **42**, 188–204 (2010).
36. E. Vig, M. Dorr, and E. Barth, "Efficient visual coding and the predictability of eye movements on natural movies," *Spatial Vis.* **22**, 397–408 (2009).
37. M. S. Mould, D. H. Foster, K. Amano, and J. P. Oakley, "A simple nonparametric method for classifying eye fixation," *Vision Res.* (submitted). doi:10.1016/j.visres.2011.12.006
38. E. Vig, Institute for Neuro- and Bioinformatics, University of Lübeck, Germany (personal communication, 2010).
39. M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," *J. Opt. Soc. Am. A* **25**, 1828–1834 (2008).
40. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," *Color Res. Appl.* **31**, 320–330 (2006).
41. CIE, Technical Committee 8-01, "A colour appearance model for colour management systems: CIECAM02," CIE 159:2004 (Commission Internationale de l'Éclairage, 2004).
42. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Res. Appl.* **26**, 340–350 (2001).
43. D. B. Montgomery and D. G. Morrison, "A note on adjusting  $R^2$ ," *J. Finance* **28**, 1009–1013 (1973).
44. A. Delorme, G. Richard, and M. Fabre-Thorpe, "Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans," *Vision Res.* **40**, 2187–2200 (2000).
45. A. Oliva and P. G. Schyns, "Diagnostic colors mediate scene recognition," *Cogn. Psychol.* **41**, 176–210 (2000).
46. D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Res.* **42**, 107–123 (2002).
47. T. J. Andrews and D. M. Coppola, "Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments," *Vision Res.* **39**, 2947–2953 (1999).
48. M. S. Castelhana and J. M. Henderson, "Stable individual differences across images in human saccadic eye movements," *Can. J. Exp. Psychol.* **62**, 1–14 (2008).
49. D. H. Foster, K. Amano, and S. M. C. Nascimento, "Color constancy in natural scenes explained by global image statistics," *Vis. Neurosci.* **23**, 341–349 (2006).
50. P. W. Lucas, N. J. Dominy, P. Riba-Hernandez, K. E. Stoner, N. Yamashita, E. Loria-Calderón, W. Petersen-Pereira, Y. Rojas-Durán, R. Salas-Pena, S. Solis-Madrugal, D. Osorio, and B. W. Darvell, "Evolution and function of routine trichromatic vision in primates," *Evolution* **57**, 2636–2643 (2003).
51. P. Sumner and J. D. Mollon, "Catarrhine photopigments are optimized for detecting targets against a foliage background," *J. Exp. Biol.* **203**, 1963–1986 (2000).
52. F. A. Wichmann, L. T. Sharpe, and K. R. Gegenfurtner, "The contributions of color to recognition memory for natural scenes," *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 509–520 (2002).
53. K. Amano, K. Uchikawa, and I. Kuriki, "Characteristics of color memory for natural scenes," *J. Opt. Soc. Am. A* **19**, 1501–1514 (2002).
54. M. S. Castelhana and J. M. Henderson, "The influence of color on the perception of scene gist," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 660–675 (2008).